

U9/331-631

#4
Cont

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification⁵ : C12N 15/29, C07K 13/00 C12N 1/21, 1/19</p>	<p>A1</p>	<p>(11) International Publication Number: WO 91/19801 (43) International Publication Date: 26 December 1991 (26.12.91)</p>
<p>(21) International Application Number: PCT/GB91/00914 (22) International Filing Date: 7 June 1991 (07.06.91) (30) Priority data: 9013016.2 11 June 1990 (11.06.90) GB (71) Applicant (for all designated States except US): MARS UK LIMITED [GB/GB]; 3D Dundee Road, Slough, Berkshire SL1 4JS (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): SPENCER, Margaret, Elizabeth [GB/GB]; 24 Totley Brook Road, Sheffield S17 3QS (GB). HODGE, Rachel [GB/GB]; 16 Mount Avenue, Leicester LE5 3RN (GB). DEAKIN, Edward, Alfred [GB/GB]; The Oakes, Oakes Park, Norton, Sheffield S8 8BA (GB). ASHTON, Sean [GB/GB]; 5 Chamberlain Court, Burncross, Chapeltown, Sheffield S30 4ZU (GB).</p>		<p>(74) Agents: SHEARD, Andrew, Gregory et al.; Kilburn & Strode, 30 John Street, London WC1N 2DD (GB). (81) Designated States: AT (European patent), AU, BE (European patent), BR, CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GB, GB (European patent), GR (European patent), HU, IT (European patent), JP, KR, LU (European patent), NL (European patent), NO, PL, RO, SE (European patent), SU, US. Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>
<p>(54) Title: RECOMBINANT 47 AND 31kD COCOA PROTEINS AND PRECURSOR (57) Abstract 47 kD and 31 kD proteins, and their 67 kD expression precursor, believed to be the source of peptide flavour precursors in cocoa (<i>Theobroma cacao</i>) have been identified. Genes coding for them have been probed, identified and sequenced, and recombinant proteins have been synthesised.</p>		

1 RECOMBINANT 47 AND 31kD COCOA PROTEINS AND PRECURSOR

2
3 This invention relates to proteins and nucleic acids derived from or otherwise
4 related to cocoa.

5
6 The beans of the cocoa plant (*Theobroma cacao*) are the raw material for cocoa,
7 chocolate and natural cocoa and chocolate flavouring. As described by Rohan
8 ("Processing of Raw Cocoa for the Market", FAO/UN (1963)), raw cocoa
9 beans are extracted from the harvested cocoa pod, from which the placenta is
10 normally removed, the beans are then "fermented" for a period of days, during
11 which the beans are killed and a purple pigment is released from the cotyledons.
12 During fermentation "unknown" compounds are formed which on roasting give
13 rise to characteristic cocoa flavour. Rohan suggests that polyphenols and
14 theobromine are implicated in the flavour precursor formation. After
15 fermentation, the beans are dried, during which time the characteristic brown
16 pigment forms, and they are then stored and shipped.

17
18 Biehl *et al*, 1982 investigated proteolysis during anaerobic cocoa seed
19 incubation and identified 26kD and 44kD proteins which accumulated during
20 seed ripening and degraded during germination. Biehl asserted that there were
21 storage proteins and suggested that they may give rise to flavour-specific
22 peptides.

23
24 Fritz *et al*, 1985 identified polypeptides of 20kD and 28kD appearing in the
25 cytoplasmic fraction of cocoa seed extracts at about 100 days after pollination.
26 It appears that the 20kD protein is thought to have glyceryl acyltransferase
27 activity.

28
29 In spite of the uncertainties in the art, as summarised above, proteins apparently
30 responsible for flavour production in cocoa beans have now been identified.
31 Further, it has been discovered that, in spite of Fritz's caution that "cocoa seed
32
33

1 *Theobroma cacao* has two primary subspecies, *Th. cacao cacao* and *Th. cacao*
2 *sphaerocarpum*. While proteins in accordance with the invention may be
3 derived from these subspecies. the invention is not limited solely to these
4 subspecies. For example, many cocoa varieties are hybrids between different
5 species; an example of such a hybrid is the trinitario variety.

6

7 The invention also relates to nucleic acid, particularly DNA, coding for the
8 proteins referred to above (whether the primary translation products, the
9 processed proteins or fragments). The invention therefore also provides, in
10 further aspects:

11

12 - nucleic acid coding for a 67kD protein of *Th. cacao*, or for a
13 fragment thereof;

14

15 - nucleic acid coding for a 47kD protein of *Th. cacao*, or for a
16 fragment thereof;

17

18 - nucleic acid coding for a 31kD protein of *Th. cacao*, or for a
19 fragment thereof;

20

21 Included in the invention is nucleic acid which is degenerate for the wild type
22 protein and which codes for conservative or other non-deleterious mutants.
23 Nucleic acid which hybridises to the wild type material is also included.

24

25 Nucleic acid within the scope of the invention will generally be recombinant
26 nucleic acid and may be in isolated form. Frequently, nucleic acid in
27 accordance with the invention will be incorporated into a vector (whether an
28 expression vector or otherwise) such as a plasmid. Suitable expression vectors
29 will contain an appropriate promoter, depending on the intended expression
30 host. For yeast, an appropriate promoter is the yeast pyruvate kinase (PK)
31 promoter: for bacteria an appropriate promoter is a strong lambda promoter.

32

33

- 1 Figure 4 shows the relationship between the 67kD protein and seed storage
2 proteins from other plants;
3
4 Figure 5 shows a map of plasmid pJLA502;
5
6 Figure 6 shows schematically the formation of plasmid pMS900;
7
8 Figure 7 shows two yeast expression vectors useful in the present invention;
9 vector A is designed for internal expression and vector B is designed for
10 secreted expression;
11
12 Figure 8a shows, in relation to vector A, part of the yeast pyruvate kinase gene
13 showing the vector A cloning site, and the use of *Hin-Nco* linkers to splice in
14 the heterologous gene;
15
16 Figure 8b shows, in relation to vector B, part of the yeast alpha-factor signal
17 sequence showing the vector B cloning site, and the use of *Hin-Nco* linkers to
18 create an in-phase fusion;
19
20 Figure 9a shows how plasmid pMS900 can be manipulated to produce plasmids
21 pMS901, pMS903, pMS907, pMS908, pMS911, pMS912 and pMS914;
22
23 Figure 9b shows how plasmid pMS903 can be manipulated to produce plasmids
24 pMS904, pMS905, pMS906, pMS909 and pMS916;
25
26 Figure 10 shows maps of plasmids pMS908, pMS914, pMS912, pMS906,
27 pMS916 and pMS910;
28
29 Figure 11 shows the construction of a plasmid to express the 67kD protein by
30 means of the AOX promoter on an integrated vector in *Hansenula polymorpha*;
31 and
32
33

1 *Characteristics of the Storage Polypeptides*

2
3 The solubility characteristics of the 47 kD and 31 kD polypeptides were roughly
4 defined by one or two quick experiments. Dialysis of the polypeptide solution
5 against SDS-free extraction buffer rendered the 47 kD and 31 kD polypeptides
6 insoluble, as judged by their ability to pass through a 0.22 micron membrane.
7 Fast Protein Liquid Chromatography (FPLC) analysis also showed that the 47
8 kD and 31 kD polypeptides were highly associated after extraction with
9 McIlvaines buffer pH 6.8 (0.2 M disodium hydrogen phosphate titrated with
10 0.1 M citric acid). The 47 kD and 31 kD polypeptides are globulins on the
11 basis on their solubility.

12

13 *Purification of the 47 kD and 31 kD polypeptides*

14

15 The 47 kD and 31 kD polypeptides were purified by two rounds of gel filtration
16 on a SUPEROSE-12 column of the PHARMACIA Fast Protein Liquid
17 Chromatography system (FPLC), or by electroelution of bands after preparative
18 electrophoresis. (The words SUPEROSE and PHARMACIA are trade marks.)
19 Concentrated protein extracts were made from 50 mg acetone powder per ml of
20 extraction buffer, and 1-2 ml loaded onto 2 mm thick SDS-PAGE gels poured
21 without a comb. After electrophoresis the gel was surface stained in aqueous
22 Coomassie Blue, and the 47 kD and 31 kD bands cut out with a scalpel. Gel
23 slices were electroeluted into dialysis bags in electrophoresis running buffer at
24 15 V for 24 hours, and the dialysate dialysed further against 0.1% SDS.
25 Samples could be concentrated by lyophilisation.

26

27 Example 2

28

29 *Amino-acid Sequence Data from Proteins*

30

31 Protein samples (about 10 µg) were subjected to conventional N-terminal
32 amino-acid sequencing. The 47 kD and 31 kD polypeptides were N-terminally
33 blocked, so cyanogen bromide peptides of the 47 kD and 31 kD peptides were

1 The gamma-globulin fraction of the serum was partially purified by
2 precipitation with 50% ammonium sulphate, solubilisation in
3 phosphate-buffered saline (PBS) and chromatography on a DE 52 cellulose
4 ion-exchange column as described by Hill, 1984. Fractions containing
5 gamma-globulin were monitored at 280 nm (OD_{280} of 1.4 is equivalent to 1
6 mg/ml gamma-globulin) and stored at -20°C .

7 The effective titre of the antibodies was measured using an enzyme-linked
8 immunosorbant assay (ELISA). The wells of a polystyrene microtitre plate
9 were coated with antigen (10-1000 ng) overnight at 4°C in carbonate coating
10 buffer. Wells were washed in PBS-Tween and the test gamma globulin added at
11 concentrations of 10, 1 and $0.1\text{ }\mu\text{g/ml}$ (approximately 1:100, 1:1000 and
12 1:10,000 dilutions). The diluent was PBS-Tween containing 2% polyvinyl
13 pyrrolidone (PVP) and 0.2% BSA. Controls were preimmune serum from the
14 same animal. Binding took place at 37°C for 3-4 hours. The wells were
15 washed as above and secondary antibody (goat anti-rabbit IgG conjugated to
16 alkaline phosphatase) added at a concentration of $1\text{ }\mu\text{g/ml}$, using the same
17 conditions as the primary antibody. The wells are again washed, and alkaline
18 phosphatase substrate (p-nitrophenyl phosphate; 0.6 mg/ml in diethanol-amine
19 buffer pH 9.8) added. The yellow colour, indicating a positive reaction, was
20 allowed to develop for 30 minutes and the reaction stopped with 3M NaOH.
21 The colour is quantified at 405 nm. More detail of this method is given in Hill,
22 1984. The method confirmed that the antibodies all had a high titre and could
23 be used at $1\text{ }\mu\text{g/ml}$ concentration.

24

25 Example 4

26

27 *Isolation of Total RNA from Immature Cocoa Beans*

28

29 The starting material for RNA which should contain a high proportion of
30 mRNA specific for the storage proteins was immature cocoa beans, at about 130
31 days after pollination. Previous work had suggested that synthesis of storage
32 proteins was approaching its height by this date (Biehl *et al.*, 1982). The beans
33 are roughly corrugated and pale pinkish-purple at this age.

1 *Preparation of mRNA From Total RNA*

2

3 The mRNA fraction was separated from total RNA by affinity chromatography
4 on a small (1 ml) oligo-dT column, the mRNA binding to the column by its
5 poly A tail. The RNA (1-2 mg) was denatured by heating at 65°C and applied
6 to the column in a high salt buffer. Poly A+ was eluted with low salt buffer,
7 and collected by ethanol precipitation. The method is essentially that of Aviv
8 and Leder (1972), modified by Maniatis *et al* (1982). From 1 mg of total
9 RNA, approximately 10-20 µg polyA+ RNA was obtained (1-2%).

10

11 *In vitro Translation of mRNA*

12

13 The ability of mRNA to support *in vitro* translation is a good indication of its
14 cleanliness and intactness. Only mRNAs with an intact polyA tail (3' end) will
15 be selected by the oligo-dT column, and only mRNAs which also have an intact
16 5' end (translational start) will translate efficiently. *In vitro* translation was
17 carried out using RNA-depleted wheat-germ lysate (Amersham International),
18 the *de novo* protein synthesis being monitored by the incorporation of [³⁵
19 S]-methionine (Roberts and Paterson, 1973). Initially the rate of *de novo*
20 synthesis was measured by the incorporation of [³⁵ S]-methionine into
21 TCA-precipitable material trapped on glass fibre filters (GFC, Whatman). The
22 actual products of translation were investigated by running on SDS-PAGE,
23 soaking the gel in fluor. drying the gel and autoradiography. The mRNA
24 preparations translated efficiently and the products covered a wide range of
25 molecular weights, showing that intact mRNAs for even the largest proteins had
26 been obtained. None of the major translation products corresponded in size to
27 the 47kD or 31kD storage polypeptides identified in mature beans, and it was
28 apparent that considerable processing of the nascent polypeptides must occur to
29 give the mature forms.

30

31

32

33

1 Example 6

2

3 *cDNA Synthesis From the mRNA Preparations*

4

5 cDNA synthesis was carried out using a kit from Amersham International. The
6 first strand of the cDNA is synthesised by the enzyme reverse transcriptase,
7 using the four nucleotide bases found in DNA (dATP, dTTP, dGTP, dCTP) and
8 an oligo-dT primer. The second strand synthesis was by the method of Gubler
9 and Hoffman (1983), whereby the RNA strand is nicked in many positions by
10 RNase H, and the remaining fragments used to prime the replacement synthesis
11 of a new DNA strand directed by the enzyme *E. coli* DNA polymerase I. Any
12 3' overhanging ends of DNA are filled in using the enzyme T4 polymerase.
13 The whole process was monitored by adding a small proportion of [³²P]-dCTP
14 into the initial nucleotide mixture, and measuring the percentage incorporation
15 of label into DNA. Assuming that cold nucleotides are incorporated at the same
16 rate, and that the four bases are incorporated equally, an estimate of the
17 synthesis of cDNA can be obtained. From 1 µg of mRNA approximately 140
18 ng of cDNA was synthesised. The products were analysed on an alkaline 1.4%
19 agarose gel as described in the Amersham methods. Globin cDNA, synthesised
20 as a control with the kit, was run on the same gel, which was dried down and
21 autoradiographed. The cocoa cDNA had a range of molecular weights, with a
22 substantial amount larger than the 600 bp of the globin cDNA.

23

24 Example 7

25

26 *Cloning of cDNA into a Plasmid Vector by Homopolymer Tailing*

27

28 The method of cloning cDNA into a plasmid vector was to 3' tail the cDNA
29 with dC residues using the enzyme terminal transferase (Boehringer Corporation
30 Ltd), and anneal into a *Pst*I-cut and 5' tailed plasmid (Maniatis *et al*, 1982
31 Eschenfeldt *et al*, 1987). The optimum length for the dC tail is 12-20 residues.
32 The tailing reaction (conditions as described by the manufacturers) was tested
33

1
2 Met-Phe-Glu-Ala-Asn-Pro
3 5' ATG TTT GAA GCT AAT CC 3'
4 C G C C
5 A
6 G

7 The actual probe was made anti-sense so that it could also be used to probe
8 mRNA. Probe synthesis was carried out using an Applied Biosystems
9 apparatus.

10
11
12 Example 9

13
14 *Use of Oligonucleotides to Probe cDNA Library*

15
16 The oligonucleotide probes were 5' end-labelled with gamma-[³²P] dATP and
17 the enzyme polynucleotide kinase (Amersham International). The method was
18 essentially that of Woods (1982, 1984), except that a smaller amount of isotope
19 (15 µCi) was used to label about 40 ng probe, in 10 mM MgCl₂, 100 mM
20 Tris-HCl, pH 7.6; 20 mM 2-mercaptoethanol.

21
22 The cDNA library was grown on GeneScreen (New England Nuclear) nylon
23 membranes placed on the surface of L-agar + 100 µg/ml ampicillin plates. (The
24 word GeneScreen is a trade mark.) Colonies were transferred from microtitre
25 plates to the membranes using a 6 x 8 multi-pronged device, designed to fit into
26 the wells of half the microtitre plate. Colonies were grown overnight at 37°C,
27 lysed in sodium hydroxide and bound to membranes as described by Woods
28 (1982, 1984). After drying the membranes were washed extensively in 3 x
29 SSC/0.1% SDS at 65°C, and hybridised to the labelled probe, using a HYBAID
30 apparatus from Hybaid Ltd, PO Box 82, Twickenham, Middlesex. (The word
31 HYBAID is a trade mark.) Conditions for hybridisation were as described by
32 Mason & Williams (1985), a T_d being calculated for each oligonucleotide
33 according to the formula:

1 replication, and the single-strands are packaged as phages extruded into the
2 medium. DNA can be prepared from these 'phages' using established methods
3 for M13 phages (Miller, 1987), and used for sequencing by the method of
4 Sanger (1977) using the reverse sequencing primer. The superinfecting phage
5 used is a derivative of M13 termed M13K07, which replicates poorly and so
6 does not compete well with the plasmid, and contains a selectable
7 kanamycin-resistance marker. Detailed methods for preparing single-strands
8 from the pTZ plasmids and helper phages are supplied by Pharmacia. DNA
9 sequence was compiled and analysed using the Staden package of programs
10 (Staden, 1986), on a PRIME 9955 computer. (The word PRIME is a trade
11 mark.)

12
13 Example 12

14
15 *Features of the 47 kD/31 kD cDNA and Deduced Amino-acid Sequence of the 67*
16 *kD Precursor*

17
18 DNA sequencing of the three positive clones, pMS600, pMS700, pMS800,
19 confirmed the overlap presumed in Figure 1. No sequence differences were
20 found in the overlapping regions (about 300 bp altogether), suggesting that the
21 three cDNAs were derived from the same gene. The sequence of the combined
22 cDNAs comprising 1818 bases is shown in Figure 2. The first ATG codon is
23 found at position 14, and is followed by an open reading frame of 566 codons.
24 There is a 104-base 3' untranslated region containing a polyadenylation signal at
25 position 1764. The oligonucleotide probe sequence is found at position 569.

26
27 The open reading frame translates to give a polypeptide of 566 amino-acids
28 (Figure 2), and a molecular weight of 65612, which is reasonably close to the
29 67 kD measured on SDS-PAGE gels. The N-terminal residues are clearly
30 hydrophobic and look like a characteristic signal sequence. Applying the rules
31 of Von Heije (1983), which predict cleavage sites for signal sequences, suggests
32 a cleavage point between amino-acids 20 and 21 (see Figure 3). The region
33 following this is highly hydrophobic and contains four Cys-X-X-X-Cys motifs.

1 Example 13

2

3 *Expression of the 67 kD Polypeptide in E. coli*

4

5 Before the 67 kD coding region could be inserted into a expression vector the
6 overlapping fragments from the three separate positive clones had to be spliced
7 into a continuous DNA segment. The method of splicing is illustrated in Figure
8 6: a *HindIII*-*BglIII* fragment from pMS600, a *BglIII*-*EcoRI* fragment from
9 pMS700 and an *EcoRI*-*SalI* fragment from pMS800 were ligated into pTZ19R
10 cut with *HindIII* and *SalI*. The resulting plasmid, containing the entire 67 kD
11 cDNA, was termed pMS900.

12

13 An *NcoI* site was introduced at the ATG start codon, using the mutagenic
14 primer:

15

16 5' TAG CAA CCA TGG TGA TCA 3'.

17

18 *In vitro* mutagenesis was carried out using a kit marketed by Amersham
19 International, which used the method of Eckstein and co-workers (Taylor *et al*,
20 1985). After annealing the mutagenic primer to single-stranded DNA the
21 second strand synthesis incorporates alpha-thio-dCTP in place of dCTP. After
22 extension and ligation to form closed circles, the plasmid is digested with *NciI*,
23 an enzyme which cannot nick DNA containing thio-dC. Thus only the original
24 strand is nicked, and subsequently digested with exonuclease III. The original
25 strand is then resynthesised, primed by the remaining DNA fragments and
26 complementing the mutated position in the original strand. Plasmids are then
27 transformed into *E. coli* and checked by plasmid mini preparations.

28

29 The 67 kD cDNA was then cloned into the *E. coli* expression plasmid, pJLA502
30 (Figure 5), on an *NcoI* - *SalI* fragment (pMS902).

31

32

33

1 mating alpha-factor downstream of the promoter, with a *HindIII* site within it to
2 create fusion proteins with incoming coding sequences. The vectors are
3 illustrated in Figure 7.

4
5 To use the vectors effectively it is desirable to introduce the foreign coding
6 region such that for vector A, the region from the *HindIII* cloning site to the
7 ATG start is the same as the yeast PK gene, and for vector B, the remainder of
8 the alpha-factor signal, including the lysine at the cleavage point. In practice
9 this situation was achieved by synthesising two sets of *HindIII* - *NcoI* linkers to
10 breach the gap between the *HindIII* cloning site in the vector and the *NcoI* at the
11 ATG start of the coding sequence. This is illustrated in Figure 8.

12
13 In order to use the yeast vector B, the hydrophobic signal sequence must first be
14 deleted from the 67 kD cDNA. Although direct evidence of the location of the
15 natural cleavage site was lacking, the algorithm of Von Heije predicts a site
16 between amino-acids 20 (alanine) and 21 (leucine). However it was decided to
17 remove amino-acids 2-19 by deletion, so that the useful *NcoI* site at the
18 translation start would be maintained.

19
20
21 For ease of construction of the yeast vectors, the strategy was to first clone the
22 *HindIII* - *NcoI* linkers into the appropriate pTZ plasmids, and then to clone the
23 linkers plus coding region into the yeast vectors on *HindIII* - *BamHI* fragments.
24 However the coding region contains an internal *BamHI* which must be removed
25 by in vitro mutagenesis, giving a new plasmid pMS903. The signal sequence
26 was deleted from pMS903 using the mutagenic primer

27
28 5' AGCATAGCAACCATGGTTGCTTTGTTCT 3'
29

30 to give pMS904. The appropriate *HindIII* - *NcoI* linkers were then cloned into
31 pMS903 and pMS904 to give pMS907 and pMS905 respectively, and the
32 *HindIII* - *BamHI* fragments (linkers + coding region) subcloned from these
33

1 concentrated 10-25 x in an AMICON mini concentrator. (The word AMICON
2 is a trade mark.) The washed cells were weighed and resuspended in lysis
3 buffer plus protease inhibitors (1 mM phenyl methyl sulphonyl fluoride
4 (PMSF); 1 µg/ml aprotinin; 0.5 µg/ml leupeptin) at a concentration of 1 g/ml.
5 1 volume acid-washed glass-beads was added and the cells broken by vortexing
6 for 8 minutes in total, in 1 minute bursts, with 1 minute intervals on ice. After
7 checking under the microscope for cell breakage, the mixture was centrifuged at
8 7000 rpm for 3 minutes to pellet the glass beads. The supernatant was removed
9 to a pre-chilled centrifuge tube, and centrifuged for 1 hour at 20,000 rpm.
10 (Small samples can be centrifuged in a microcentrifuge in the cold.) The
11 supernatant constitutes the soluble fraction. The pellet was resuspended in 1 ml
12 lysis buffer plus 10% SDS and 1% mercaptoethanol and heated at 90°C for 10
13 minutes. After centrifuging for 15 minutes in a microcentrifuge the supernatant
14 constitutes the particulate fraction.

15

16 Samples of each fraction and the concentrated medium were examined by
17 Western blotting. Considering first the plasmids designed for internal
18 expression in YVA, pMS908 produced immunoreactive proteins at 67 kD and
19 16 kD within the cells only. There was no evidence of the 67 kD protein being
20 secreted under the influence of its own signal sequence. The smaller protein is
21 presumed to be a degradation product. A similar result, but with improved
22 expression, was obtained with pMS914, in which the plant terminator is
23 replaced by a yeast terminator. However in pMS912, in which the coding
24 region for the hydrophilic domain has been deleted, no synthesis of
25 immunoreactive protein occurred.

26

27 For industrial production of heterologous proteins in yeast a secreted mode is
28 preferable because yeast cells are very difficult to disrupt, and downstream
29 processing from total cell protein is not easy. The results from the vectors
30 constructed for secreted expressed were rather complicated. From the simplest
31 construct, pMS906, in which the yeast α-factor signal sequence replaces the
32 plant protein's own signal, immunoreactive proteins of approximately 47 kD, 28
33 kD and 18-20 kD were obtained and secreted into the medium. At first sight

1 promoter is completely repressed. This means that cells containing the
 2 heterologous gene can be grown to a high density on glucose, and induced to
 3 produce the foreign protein by allowing the glucose to run out and adding
 4 methanol.

5
 6 A plasmid, pHGL1, containing the MOX promoter and terminator, and a
 7 cassette containing the yeast α -factor secretory signal sequence, were prepared.
 8 The 67 kD coding region was cloned into pHGL1 on a *Bam*HI - *Bam*HI
 9 fragment, replacing the *Bgl*III fragment which contains the 3' end of the MOX
 10 coding region. The whole promoter - gene - terminator region can then be
 11 transferred to YEp13 on a *Bam*HI - *Bam*HI fragment to give the expression
 12 plasmid pMS922. The details of the construction are illustrated in Figure 11.
 13 An analogous expression plasmid, pMS925, has been constructed with the yeast
 14 α -factor spliced onto the 67 kD coding region, replacing the natural plant
 15 signal. The *Bam*HI - *Hind*III cassette containing the α -factor was ligated to the
 16 *Hind*III - *Bam*HI fragment used to introduce the 67 kD coding region into YVB.
 17 The α -factor plus coding region was then cloned with pHGL1 on a *Bam*HI -
 18 *Bam*HI fragment, and transferred into YEP13 as before. Details are shown in
 19 Figure 12.

20
 21 Both constructs have been transformed into *Hansenula* and grown under
 22 inducing conditions with 0.5% or 1% methanol. Both constructs directed the
 23 production of immunoreactive protein within the cells, and pMS925 secreted the
 24 protein into the medium under the influence of the α -factor signal sequence.

25 26 *E. coli* Strains

27
 28 RR1 $F^{-}v_B^{-}M_B$ *ara-14 proA2 leuB6 lacY1 galK2 vpsL20 (str^r)*
 29 *xyl-5 ml-1 supE44*

30
 31 CAG629 *lac_{am} rvp_{am} pho_{am} htpR_{am} mal rpsL lon supC_{ts}*

- 1 Cuming, A.C., Williams, R.S., and Cullimore, J.V. in "Immunology in Plant
2 Science", Ed. Wang, T.L., Cambridge University Press, 1986. The use of
3 Antibodies in Molecular Biology.
4
- 5 Doyle, J.J., Schuler, M.A., Godette, W.D., Zencer, V., Beachy, R.N., and
6 Slightom, J.L. *J. Biol. Chem.* 261, 9228-9238 (1986). The glycosylated seed
7 storage proteins of Glycine max and Phaseolus vulgaris : Structural homologies
8 of genes and proteins.
9
- 10 Dretzen, G., Bellard, M., Sassone-Corsi, P., and Chambon, P. *Analytical*
11 *Biochemistry* 112, 295-298 (1981). A reliable method for the recovery of DNA
12 fragments from agarose and acrylamide gels.
13
- 14 Eschenfeldt, W.H., Puskas, R.S., and Berger, S.L. *Methods in Enzymology*
15 152, 337-342 (1987). Homopolymeric Tailing.
16
- 17 Fritz *et al* (*J. Food Sci.* 50 946-950 (1985))
18
- 19 Gubler, U., and Hoffman, B.J. *Gene* 25, 263 (1983). A simple and very
20 efficient method for generating cDNA libraries.
21
- 22 Hall, T.C., Ma, Y., Buchbinder, B.U., Pyne J.W., Sun, S.M., and Bliss,
23 F.A. *Proc. Natl. Acad. Sci. USA* 75, 3196-3200 (1978). Messenger RNA for
24 G1 protein of French bean seeds : cell-free translation and product
25 characterisation.
26
- 27 Hill, S.A. "Methods in Plant Virology", Blackwell 1984.
28
- 29 Johnston, J.R. in "Yeast : A practical approach". Eds Campbell, I., and
30 Duffus, J.H. IRL Press. 1988. Yeast Genetics, Molecular Aspects.
31
- 32 Kreil, G. *Annual Rev. Biochem.* 50, 317-348 (1981). Transfer of proteins
33 across membranes.

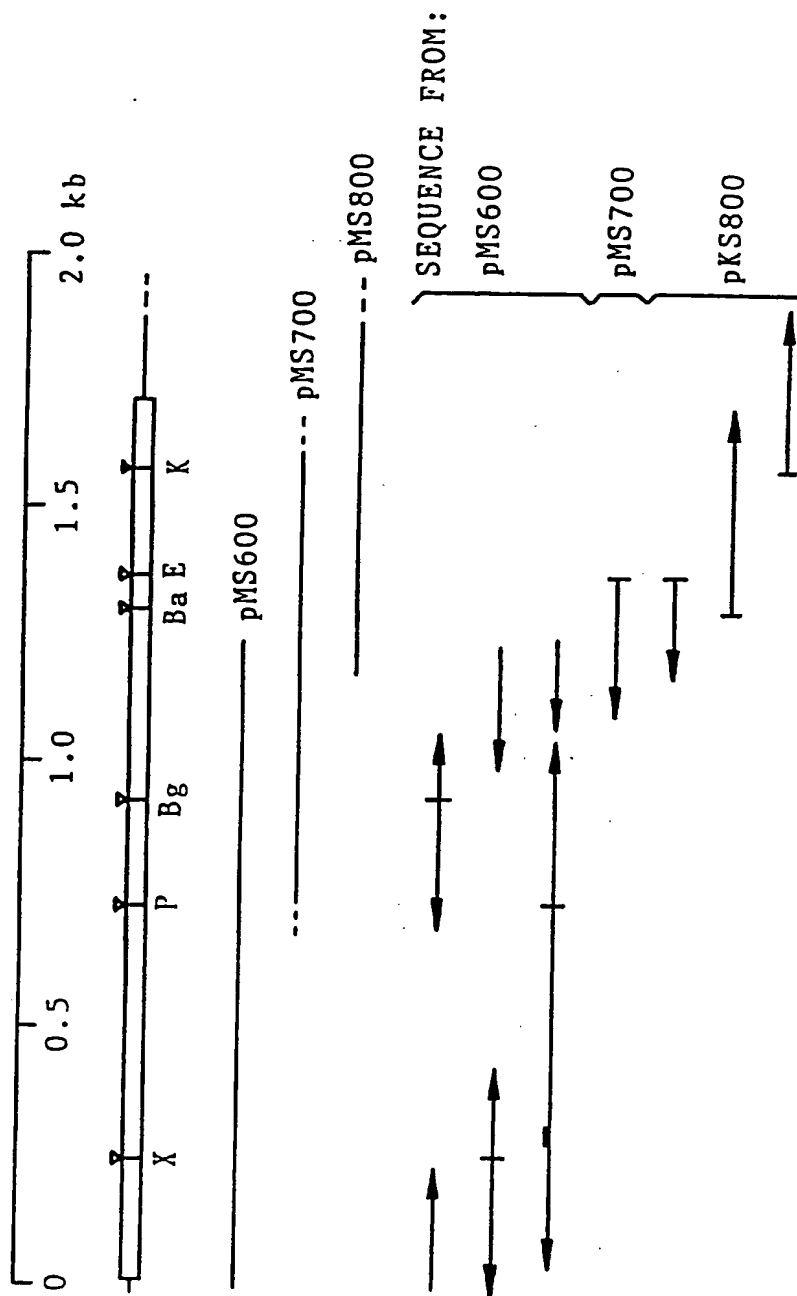
- 1 Roberts, B.E., and Paterson, B.M. *Proc. Natl. Acad. Sci. USA* 70, 2330
2 (1973). Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S
3 RNA in a cell-free system from commercial wheat germ.
4
- 5 Sanger, F., Nicklen, S., and Coulson, A.R. *Proc. Natl. Acad. Sci. USA* 74,
6 5463-5467 (1977). DNA sequencing with chain-terminating inhibitors.
7
- 8 Schauder, B., Blocker, H., Frank, R., and McCarthy, J.E.G. *Gene* 52,
9 279-283 (1987). Inducible expression vectors incorporating the *E. coli* *ap_rE*
10 translation initiation region.
11
- 12 Scott, R., Draper, J., Jefferson, R., Dury, G., and Jacob, L. in "Plant Genetic
13 Transformation and Gene Expression: A Laboratory Manual". Eds. Draper,
14 J., Scott, R., Armitage, P., Walden, R. Blackwell 1988. Analysis of gene
15 organisation and expression in plants.
16
- 17 Staden, R. *Nucleic Acids Res.* 14, 217-231 (1986). The current status and
18 portability of our sequence handling software.
19
- 20 Sudbery, P.E., Gleeson, M.A., Veale, R.A., Lederboer, A.M., and
21 Zoetmulder, M.C.M. *Biocem. Soc. Trans* 16 1081-103 (1988). Hansenula
22 polymorpha as a novel yeast system for the expression of heterologous genes.
23
- 24 Taylor, J.W., Ott, J., and Eckstein, F. *Nucleic Acids Res.* 13, 8765-8785
25 (1985). The rapid generation of oligonucleotide-directed mutations at high
26 frequency using phosphorothioate-modified DNA.
27
- 28 Towbin, H., Staehelin, T., and Gordon, J. *Proc. Natl. Acad. Sci. USA* 76,
29 4350-4534 (1979). Electrophoretic transfer of proteins from polyacrylamide
30 gels to nitrocellulose sheets : procedure and some applications.
31
- 32 Von Heije, G. *Eur. J. Biochem* 133, 17-21 (1983). Patterns of Amino-acids
33 near Signal-Sequence Cleavage Sites.

CLAIMS

- 1
- 2
- 3 1. A 67kD protein of *Theobroma cacao*, or a fragment thereof.
- 4
- 5 2. A 47kD protein of *Th. cacao*, or a fragment thereof.
- 6
- 7 3. A 31kD protein of *Th. cacao*, or a fragment thereof.
- 8
- 9 4. A protein as claimed in claim 1, 2 or 3, having at least part of the
- 10 sequence shown in Figure 2.
- 11
- 12 5. A fragment as claimed in any one of claims 1 to 4, which comprises at
- 13 least four amino acids.
- 14
- 15 6. A protein or fragment as claimed in any one of claims 1 to 6, which is
- 16 recombinant.
- 17
- 18 7. Recombinant or isolated nucleic acid coding for a protein or fragment as
- 19 claimed in any one of claims 1 to 5.
- 20
- 21 8. Nucleic acid as claimed in claim 7 which is DNA.
- 22
- 23 9. Nucleic acid as claimed in claim 8, having at least part of the sequence
- 24 shown in Figure 2.
- 25
- 26 10. Nucleic acid as claimed in claim 7, 8 or 9, which is in the form of a
- 27 vector.
- 28
- 29 11. Nucleic acid as claimed in claim 10, wherein the vector is an expression
- 30 vector and the protein- or fragment-coding sequence is operably linked to a
- 31 promoter.
- 32
- 33

1/22

FIG. 1



SUBSTITUTE SHEET

3/22

R D E E G N F K I L Q R F A E N S P P L
CAGGGATGAAGAGGGCAACTTCAAGATCCTCCAGAGGTTTGCTGAGAAGTCTCCTCCACT
490 500 510 520 530 540

K G I N D Y R L A M F E A N P N T F I L
CAAGGGCATCAACGATTACCGCTTGCCATGTTGGAAGCAAAATCCCAACACTTTTATTCT
550 560 570 580 590 600

P H H C D A E A I Y F V T N G K G T I T
TCCGCACCACTGTGATGCTGAGGCAATTACTTCGTGACAAACGGAAGGGGACAATTAC
610 620 630 640 650 660

F V T H E N K E S Y N V Q R G T V V S V
GTTTGTGACTCATGAAACAAAGAGTCCTATAATGTACAGCGTGGAAACAGTAGTCAGCGT
670 680 690 700 710 720

P A G S T V Y V V S Q D N Q E K L T I A
TCCTGCAGGAAGCACTGTTTACGTGGTTAGCCAAAGACAACCAAGAGAAGCTAACCATAGC
730 740 750 760 770 780

V L A L P V N S P G K Y E L F F P A G N
TGTGCTGCCCCCTGTTAATTCCTCCTGGCAAATATGAGTTATTCTCCCGCTGGAAA
790 800 810 820 830 840

N K P E S Y Y G A F S Y E V L E T V F N
TAATAAACCTGAATCATATACGGAGCCTTCAGCTATGAAGTTCTTGAGACCGCTCTTCAA
850 860 870 880 890 900

T Q R E K L E E I L E E Q R G Q K R Q Q
TACACAAAGAGAGAAGCTGGAGGAGATCTTGGAGGAACAGAGAGGGCAGAGAGGCAGCA
910 920 930 940 950 960

FIG. 2B

SUBSTITUTE SHEET

5/22

A S K D Q P L N A V A F G L N A Q N N Q
TGCAATCAAAGACCAAGCCCTGAATGCAGTTGCGTTTGGACTCAACGCCCAAGAACCA
1450 1460 1470 1480 1490 1500

R I F L A G K K N L V R Q M D S E A K E
GAGAATTTTCCTTGCAGGGAAGAAAGAACTTGGTCAGACAAATGGATAGCGAGGCAAGGA
1510 1520 1530 1540 1550 1560

L S F G V P S K L V D N I F N N P D E S
GTTATCATTTGGGGTACCATCGAAATGCTAGATAATATATCAACAACCCGGATGAGTC
1570 1580 1590 1600 1610 1620

Y F M S F S Q Q R R R D E R R G N P L
GTATTTCAATGTCTTTCTCTCAACAGAGGCGCTCGAGATGAAGAGGGGCAATCCCTT
1630 1640 1650 1660 1670 1680

A S I L D F A R L F *
GGCTCAATTTGGACTTTGCCCGCTTGTCTAAGCAGCTGCTTCCACTTTTGTATCAGA
1690 1700 1710 1720 1730 1740

CATGCAGAGGCATGTAATGCAATAAATAAGTTGGCCTATGTAAAGAGAGAGTTTGCT
1750 1760 1770 1780 1790 1800

TTTGTCTTGTCTAACCTTGTGTTTGAAGTAACTTCAATGTAAATGAGAGTTGTAT
1810 1820 1830 1840 1850 1860

CTTTCTA

FIG. 2D

SUBSTITUTE SHEET

9 / 22

FIG. 4C

[illegible]

SUBSTITUTE SHEET

11/22

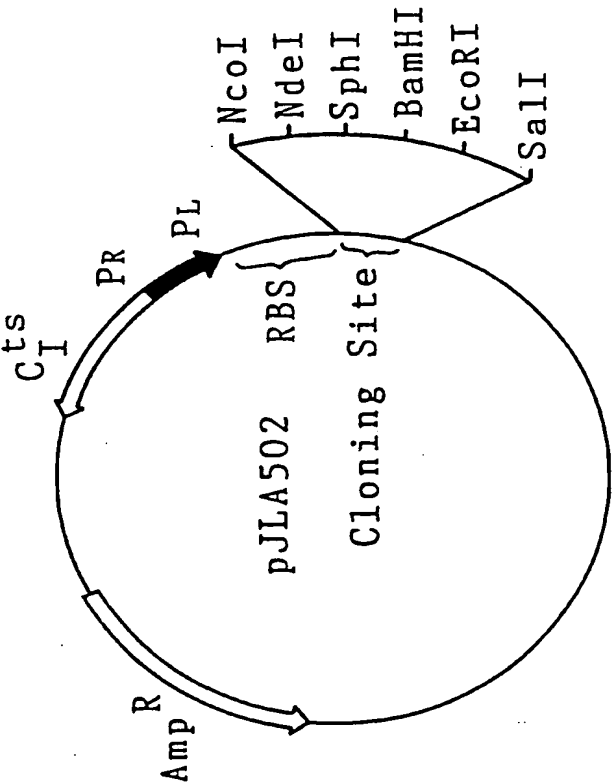
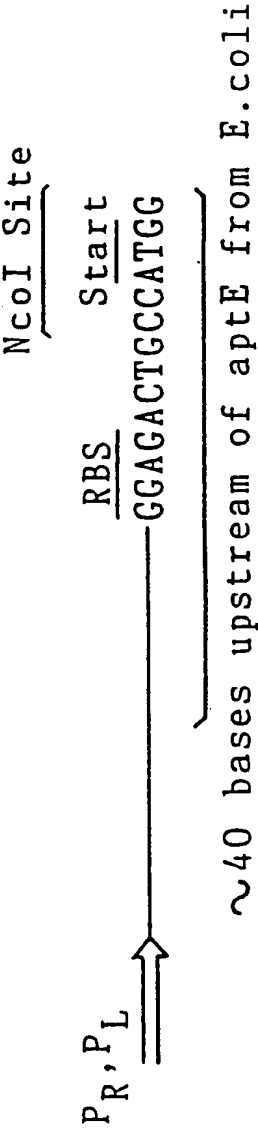


FIG. 5



SUBSTITUTE SHEET

13/22

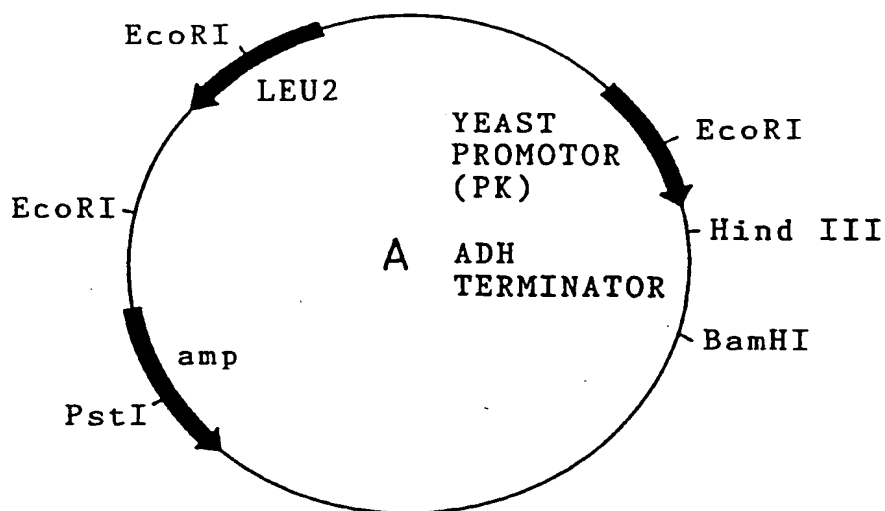
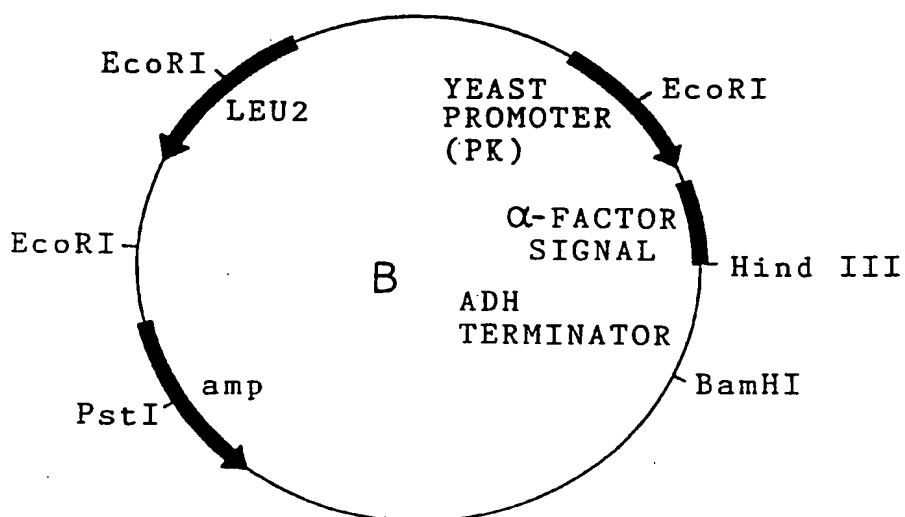


FIG. 7



SUBSTITUTE SHEET

FIG. 8B

1 231
 Met-----GluGlyValSerLeuAspLysArgGlu
 ATG-----GAAGGGGTAAAGCTTGGATAAAAGAGAG
 Hin

YEAST ALPHA-FACTOR
 SIGNAL SEQUENCE

15/22

AGCTTGGATAAAAGAGC
 ACCTATTTTCTCGGTAC

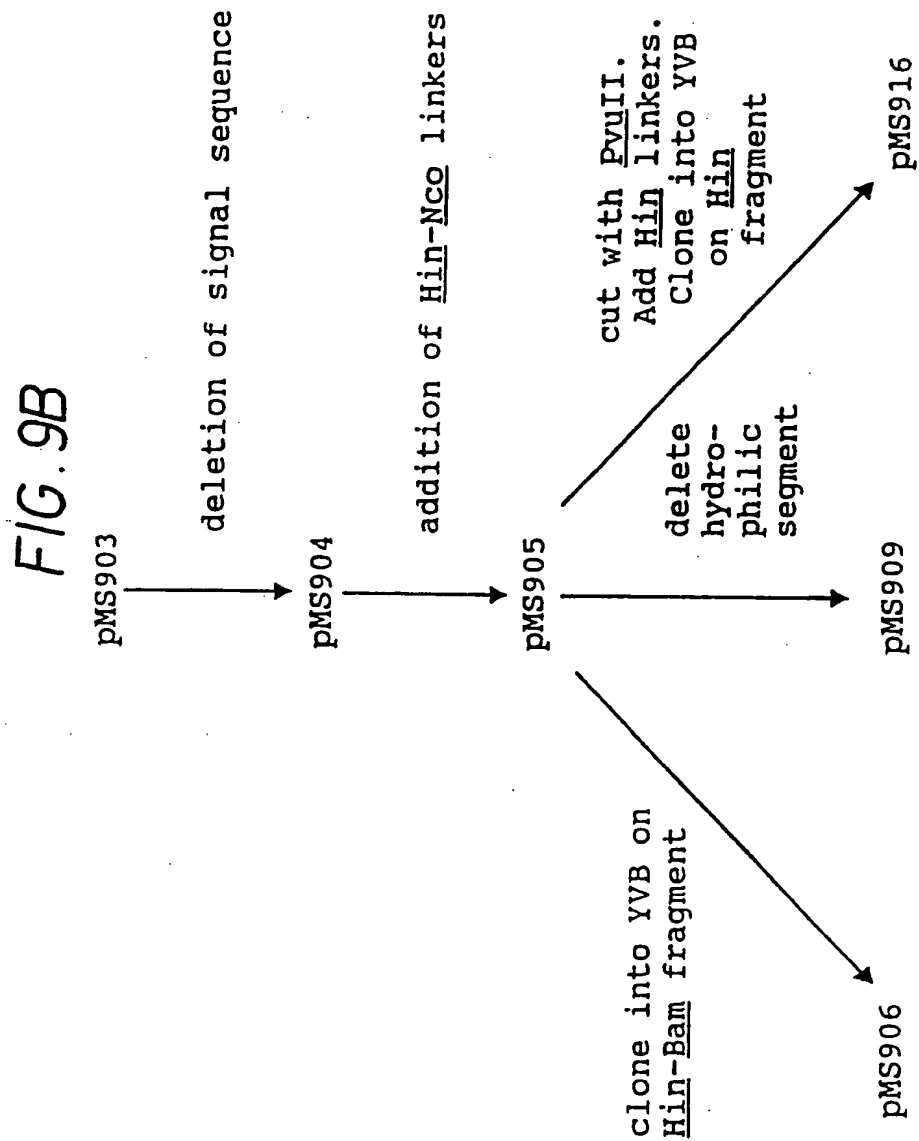
HIN-NCO LINKERS

Met---GluGlyValSerLeuAspLysArgAlaMetAlaLeu
 ATG---GAAGGGGTAAAGCTTGGATAAAAGAGCCATGGCGTTG
 Hin Nco

IN-PHASE FUSION OF
 67KD CODING REGION

SUBSTITUTE SHEET

17/22



SUBSTITUTE SHEET

19/22

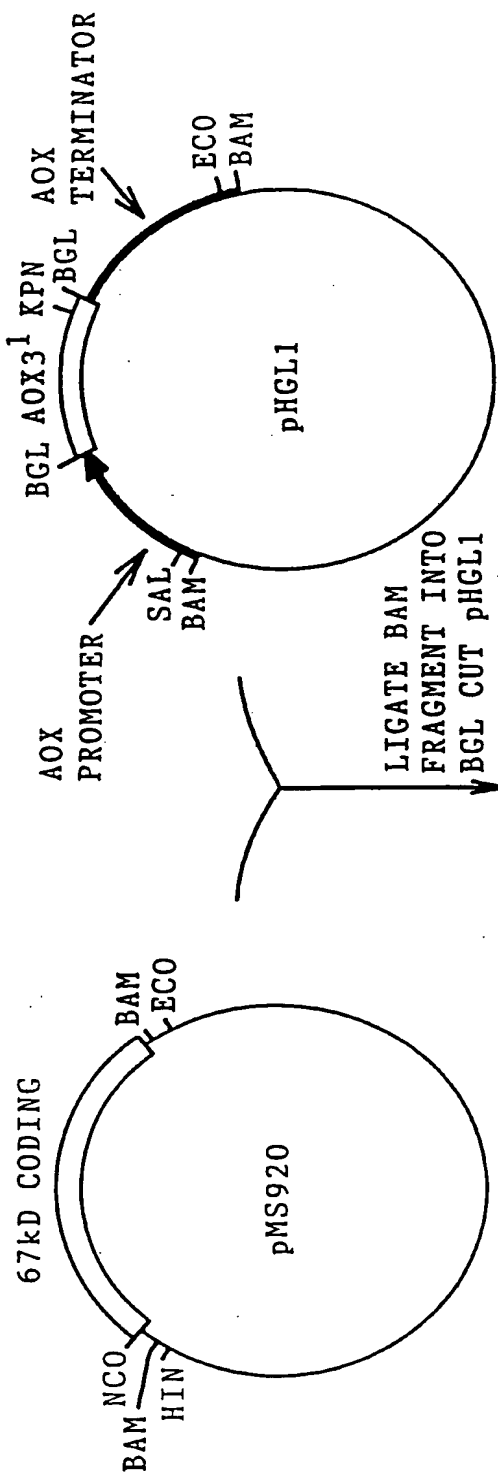
HIN NCO 67kd CODING HIN
HIN-HIN FRAGMENT
FROM pMS914

FILL IN OVER HANGING ENDS.
LIGATE ON BAM LINKERS.
CUT WITH BAM

BAM NCO 67kd CODING BAM

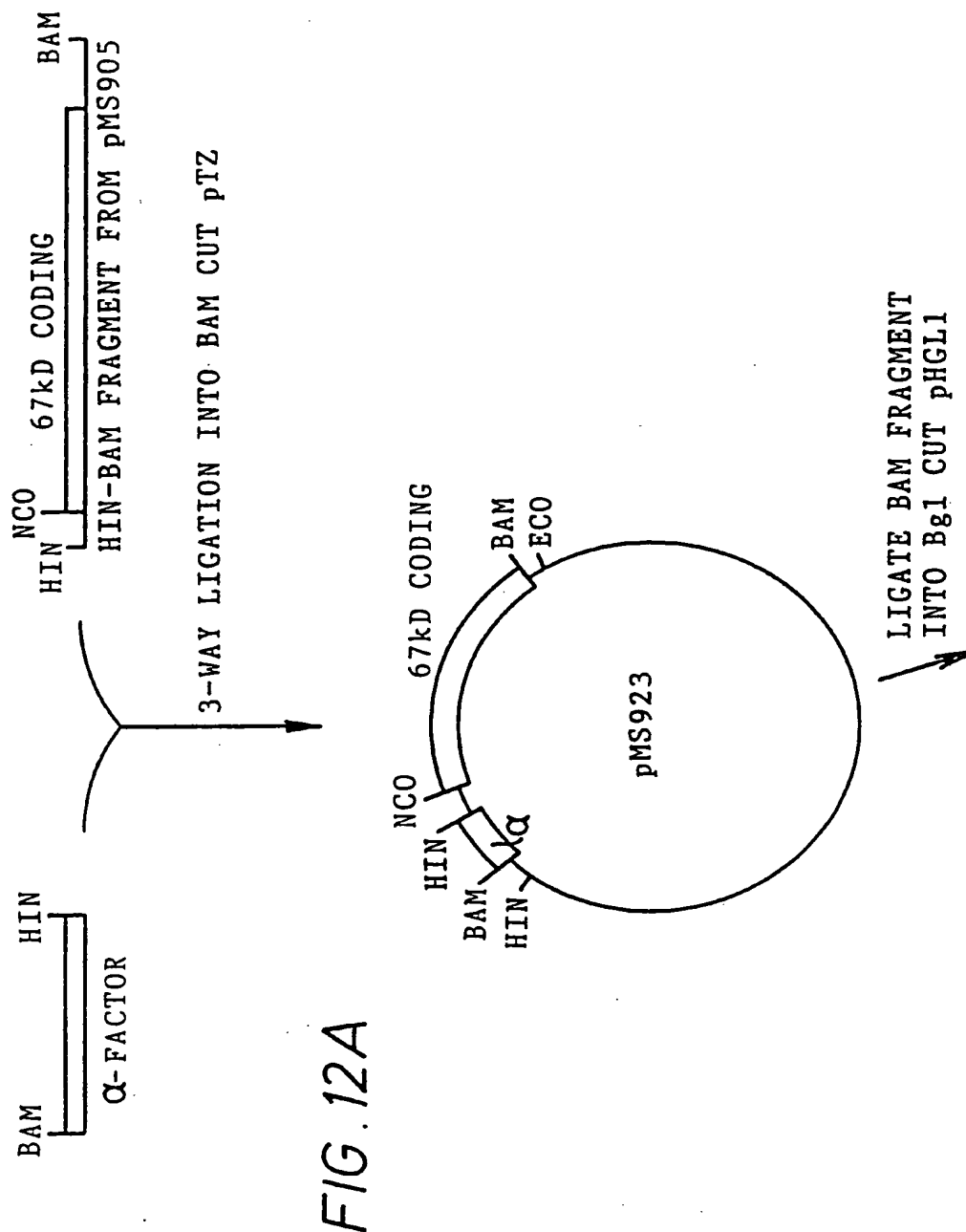
FIG. 11A

LIGATE INTO BAM CUT
pT2



SUBSTITUTE SHEET

21/22




SUBSTITUTE SHEET

INTERNATIONAL SEARCH REPORT

PCT/GB 91/00914

International Application No

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) ⁶		
According to International Patent Classification (IPC) or to both National Classification and IPC		
Int.Cl. 5 C12N15/29 ; C07K13/00 ; C12N1/21 ; C12N1/19		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁷		
Classification System	Classification Symbols	
Int.Cl. 5	C07K ; C12N	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched ⁸		
III. DOCUMENTS CONSIDERED TO BE RELEVANT⁹		
Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
X	J. SCI. FOOD AGRIC. vol. 33, 1982, pages 1291 - 1304; BIEHL, B., ET AL: 'Vacuolar storage proteins of cocoa seeds and their degradation during germination and fermentation' see the whole document ---	1-6
X	J. FOOD SCIENCE vol. 50, 1985, pages 946 - 950; FRITZ, P. J., ET AL: 'Cocoa seeds: Changes in protein and polysomal RNA during development' see the whole document --- -/-	1-6
<p>¹⁰ Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"A" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report	
30 SEPTEMBER 1991	20. 11. 91	
International Searching Authority	Signature of Authorized Officer	
EUROPEAN PATENT OFFICE	MADDOX A.D. 	

Form PCT/ISA/210 (second sheet) (January 1985)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
X	PLANT MOL BIOL vol. 11, 1988, pages 683 - 695; HIGGINS T.J.V., ET.AL.: 'The sequence of a pea vicilin gene and its expression in transgenic tobacco plants ' see page 684 methods section ---	4-11, 15, 17, 19

Form PCT/ISA/210 (contd sheet) (January 1983)